



(12) 发明专利

(10) 授权公告号 CN 109964270 B

(45) 授权公告日 2020.09.25

(21) 申请号 201780025827.X

(72) 发明人 周荣

(22) 申请日 2017.10.24

(74) 专利代理机构 成都七星天知识产权代理有限公司 51253

(65) 同一申请的已公布的文献号  
申请公布号 CN 109964270 A

代理人 杨永梅

(43) 申请公布日 2019.07.02

(51) Int.Cl.

(85) PCT国际申请进入国家阶段日  
2018.11.05

G10L 15/02 (2006.01)

(86) PCT国际申请的申请数据  
PCT/CN2017/107432 2017.10.24

审查员 陈成

(87) PCT国际申请的公布数据  
W02019/079957 EN 2019.05.02

(73) 专利权人 北京嘀嘀无限科技发展有限公司  
地址 100193 北京市海淀区东北旺路西路8  
号院34号楼

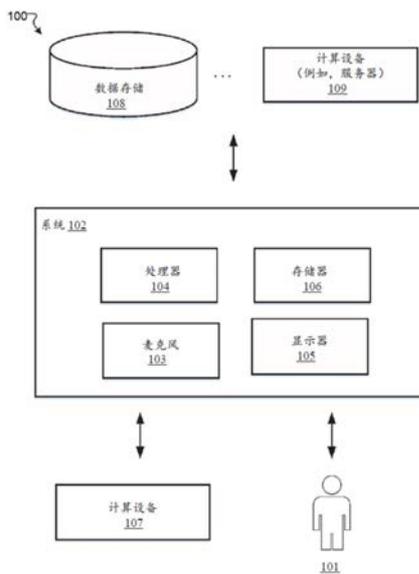
权利要求书3页 说明书13页 附图7页

(54) 发明名称

用于关键短语识别的系统和方法

(57) 摘要

用于关键短语识别的方法可以包括:获得音频;获得对应于至少两个音频部分的至少两个候选词,并获得所获得的候选词与音频部分之间的每个对应关系的第一概率分数;确定至少两个候选词是否分别匹配关键短语的至少两个关键词,以及至少两个候选词的每个的第一概率分数是否超过对应的第一阈值,至少两个候选词构成一个候选短语;以及响应于确定至少两个候选词匹配至少两个关键词且每个第一概率得分超过对应的阈值,基于至少两个候选词的每个的第一概率得分,获取表示候选短语与关键短语之间的匹配关系的第二概率得分。



1. 一种关键短语识别的方法,包括:

获得包含音频部分的序列的音频;

获得对应于至少两个所述音频部分的至少两个候选词,并获得所述获得的候选词与所述音频部分之间的每个对应关系的第一概率分数;

确定所述至少两个候选词是否分别匹配关键短语的至少两个关键词,以及所述至少两个候选词的每个的第一概率分数是否超过对应的第一阈值,所述至少两个候选词构成候选短语;

响应于确定所述至少两个候选词匹配所述至少两个关键词且每个所述第一概率分数超过所述对应的阈值,基于所述至少两个候选词的每个的第一概率分数,获得表示所述候选短语与关键短语之间的匹配关系的第二概率得分;以及

响应于确定所述第二概率得分超过第二阈值,将所述候选短语确定为所述关键短语。

2. 根据权利要求1所述的方法,其特征在于:

获得对应于所述至少两个所述音频部分的所述至少两个候选词,并获得所述获得的候选词与所述音频部分之间每个对应关系的第一概率分数,包括:获得与所述音频对应的声谱图;

沿着所述声谱图获得每个时间帧的特征向量,以获得对应于所述声谱图的至少两个特征向量;

获得对应于所述至少两个特征向量的至少两个语言单元;

至少基于将语言单元映射到单词的词典获得对应于所述音频的候选词序列,并且对于所述每个候选词,至少基于用语言单元的样本序列训练的模型获得所述第一概率分数;以及

从所述候选词序列中获取所述至少两个候选词。

3. 根据权利要求2所述的方法,其特征在于,所述方法进一步包括:

至少基于所述时间帧确定所述获得的音频中的所述关键短语的开始时间和结束时间。

4. 根据权利要求1所述的方法,其特征在于:

所述至少两个候选词是按时间顺序;以及

所述至少两个候选词和所述至少两个关键词之间的各自匹配包括所述候选短语中的顺序次序中的候选词与所述关键短语中的相同顺序次序中的关键词之间的匹配。

5. 根据权利要求4所述的方法,其特征在于:

确定所述至少两个候选词是否分别匹配所述至少两个所述关键短语的关键词,以及所述至少两个候选词的每个的第一概率分数是否超过对应的第一阈值,包括:

分别在前向或后向顺序次序中,确定所述至少两个候选词和所述至少两个关键词之间的各自匹配。

6. 根据权利要求1所述的方法,其特征在于,所述方法进一步包括:

响应于确定所述至少两个候选词中的任何一个的第一概率分数不超过对应阈值,不将所述候选短语确定为所述关键短语。

7. 根据权利要求1所述的方法,其特征在于:

所述方法不是基于或部分基于语音模型实现的;以及

所述方法不是由语音解码器实现或部分由语音解码器实现的。

8. 根据权利要求1所述的方法,其特征在于:  
所述关键短语包括唤醒应用程序的短语、标准化语言的短语,或紧急触发短语中的至少一个。
9. 根据权利要求1所述的方法,其特征在于:  
所述方法可由包括麦克风的移动设备实现;以及  
所述获得的音频包括由车辆中的一个或以上乘员的麦克风记录的语音。
10. 一种关键短语识别系统,包括:  
一个处理器;和  
一种存储指令的非暂时性计算机可读存储介质,当该指令由所述处理器执行时,使所述处理器执行方法,所述方法包括:  
获得包含音频部分的序列的音频;  
获得对应于至少两个所述音频部分的至少两个候选词,并获得所述获得的候选词与所述音频部分之间的每个对应关系的第一概率分数;  
确定所述至少两个候选词是否分别匹配关键短语的至少两个关键词,以及所述至少两个候选词的每个的第一概率分数是否超过对应的第一阈值,所述至少两个候选词构成候选短语;  
响应于确定所述至少两个候选词匹配所述至少两个关键词和所述每个第一概率分数超过所述对应的阈值,基于所述至少两个候选词的每个的第一概率分数,获得表示所述候选短语与关键短语之间的匹配关系的第二概率得分;以及  
响应于确定所述第二概率得分超过第二阈值,将所述候选短语确定为所述关键短语。
11. 根据权利要求10所述的系统,其特征在于:  
获得对应于所述至少两个所述音频部分的所述至少两个候选词,并获得所述获得的候选词与所述音频部分之间每个对应关系的第一概率分数,包括:  
获得与所述音频对应的声谱图;  
沿着所述声谱图获得每个时间帧的特征向量,以获得对应于所述声谱图的至少两个特征向量;  
获得对应于所述至少两个特征向量的至少两个语言单元;  
至少基于将语言单元映射到单词的词典获得对应于所述音频的候选词序列,并且对于所述每个候选词,至少基于用语言单元的样本序列训练的模型获得所述第一概率分数;以及  
从所述候选词序列中获取所述至少两个候选词。
12. 根据权利要求11所述的系统,还使得所述处理器执行:  
至少基于所述时间帧确定所述获得的音频中的所述关键短语的开始时间和结束时间。
13. 根据权利要求10所述的系统,其特征在于:  
所述至少两个候选词是按时间顺序;以及  
所述至少两个候选词和所述至少两个关键词之间的分别匹配包括所述候选短语中的顺序次序中的候选词与所述关键短语中的相同顺序次序中的关键词之间的匹配。
14. 根据权利要求13所述的系统,其特征在于:  
确定所述至少两个候选词是否分别匹配所述至少两个所述关键短语的关键词,以及所

述至少两个候选词的每个的第一概率分数是否超过对应的第一阈值,所述处理器执行:

在前向或后向顺序次序中,确定所述至少两个候选词和所述至少两个关键词之间的分别匹配。

15. 根据权利要求10所述的系统,其中,进一步使所述处理器执行:

响应于确定所述至少两个候选词中的任何一个的第一概率分数不超过对应阈值,不将所述候选短语确定为所述关键短语。

16. 根据权利要求10所述的系统,其特征在于:

不使所述处理器实现语音模型;以及

不使所述处理器实现语音解码器。

17. 根据权利要求10所述的系统,其特征在于:

所述关键短语包括唤醒应用程序的短语、标准化语言的短语或紧急触发短语中的至少一个。

18. 根据权利要求10所述的系统,还包括:

麦克风被配置为接收所述音频并将所述获得的音频发送到所述处理器,其中:

所述系统可在移动设备上实现,所述移动设备包括移动电话;以及

所述获得的音频包括车辆中的一个或以上乘员的讲话。

19. 一种用于关键短语识别的非暂时性计算机可读介质,包括存储在其中的指令,其中,所述指令在由一个或以上处理器执行时使所述一个或以上处理器执行包括以下的方法:

获得包含音频部分的序列的音频;

获得对应于至少两个所述音频部分的至少两个候选词,并获得所述获得的候选词与所述音频部分之间的每个对应关系的第一概率分数;

确定所述至少两个候选词是否分别匹配关键短语的至少两个关键词,以及所述至少两个候选词的每个的第一概率分数是否超过对应的第一阈值,所述至少两个候选词构成候选短语;

响应于确定所述至少两个候选词匹配所述至少两个关键词和所述每个第一概率分数超过所述对应的阈值,基于所述至少两个候选词的每个的第一概率分数,获得表示所述候选短语与关键短语之间的匹配关系的第二概率得分;以及响应于确定所述第二概率得分超过第二阈值,将所述候选短语确定为所述关键短语。

20. 根据权利要求19所述的非暂时性计算机可读介质,其中:

获得对应于所述至少两个所述音频部分的所述至少两个候选词,并获得所述获得的候选词与所述音频部分之间的每个对应关系的第一概率分数;包括:

获得与所述音频对应的声谱图;

沿着所述声谱图获得每个时间帧的特征向量,以获得对应于所述声谱图的至少两个特征向量;

获得对应于所述至少两个特征向量的至少两个语言单元;

至少基于将语言单元映射到单词的词典获得对应于所述音频的候选词序列,并且对于所述每个候选词,至少基于用语言单元的样本序列训练的模型获得所述第一概率分数;以及从所述候选词序列中获取所述至少两个候选词。

## 用于关键短语识别的系统和方法

### 技术领域

[0001] 本公开一般涉及用于语音识别中的关键词语识别的方法和技术。

### 背景技术

[0002] 人机交互的进步可以允许人们使用他们的声音来实现对机器的控制。例如,可以通过语音输入来实现基于诸如键盘、鼠标或触摸屏的传统指令输入的命令触发。然而,为了简化该过程,仍有许多障碍需要克服。

### 发明内容

[0003] 本申请的各种实施例包括用于关键短语识别的系统、方法和非暂时性计算机可读介质。用于关键短语识别的示例性方法可以包括:获得包括音频部分的序列的音频;获得对应于至少两个音频部分的至少两个候选词,并获得所获得的候选词与音频部分之间的每个对应关系的第一概率分数;确定至少两个候选词是否分别匹配关键短语的至少两个关键词,以及至少两个候选词的每个的第一概率分数是否超过对应的第一阈值,至少两个候选词构成候选短语;响应于确定至少两个候选词匹配至少两个关键词且每个第一概率得分超过对应的阈值,基于至少两个候选词的每个的第一概率得分,获取表示候选短语与关键短语之间的匹配关系的第二概率得分;以及响应于确定第二概率分数超过第二阈值,将候选短语确定为关键短语。

[0004] 在一些实施例中,该方法可以由包括麦克风、处理器和存储指令的非暂时性计算机可读存储介质的移动设备实现。例如,麦克风可以被配置用于接收音频,并且当指令由处理器执行时,使处理器执行该方法。所获得的音频可以包括由车辆中的一个或以上乘员的麦克风记录的语音。移动设备可以包括移动电话。

[0005] 在一些实施例中,获得对应于音频部分的至少两个候选词,并获得所获得的候选词与音频部分之间的每个对应关系的第一概率分数可包括获得对应于所述音频的声谱图,沿着所述声谱图获得每个时间帧的特征向量,以获得对应于声谱图的至少两个特征向量,获得对应于至少两个特征向量的至少两个语言单元,至少基于将语言单元映射到单词的词典获得对应于所述音频的候选词序列,并且对于所述每个候选词,至少基于用语言单元的样本序列训练的模型获得所述第一概率分数,并从候选词序列中获取至少两个候选词。

[0006] 在一些实施例中,该方法可以进一步包括至少基于时间帧确定所获得的音频中的关键短语的开始时间和结束时间。

[0007] 在一些实施例中,至少两个候选词可以是按时间顺序(即,从候选词序列和相同词序列中获得的连续词),并且至少两个候选词和至少两个关键字之间的分别匹配可以包括候选短语中的顺序次序中的候选词与关键短语中的相同顺序次序中的关键词之间的匹配。

[0008] 在一些实施例中,确定至少两个候选词是否分别匹配关键短语的至少两个关键词,以及至少两个候选词的每个的第一概率分数是否超过对应的第一阈值,则包括在前向或后向顺序次序中确定至少两个候选词和至少两个关键词之间的分别匹配。

[0009] 在一些实施例中,该方法可以进一步包括:响应于确定所述至少两个候选词中的任何一个的第一概率分数不超过对应阈值,不将所述候选短语确定为所述关键短语。

[0010] 在一些实施例中,该方法可以不基于或部分地基于语音模型来实现,并且该方法可以不由语音解码器实现或部分地由语音解码器实现。

[0011] 在一些实施例中,该方法可以被实现为从音频中发现至少两个关键短语,并且所述至少两个关键短语可以包括用于唤醒应用程序的短语,标准化语言的短语或者紧急触发短语中的至少一个。

[0012] 在考虑了组成说明书一部分的以下描述、所附权利要求书与附图后,本申请公开的系统、方法和非暂时性计算机可读介质的这些和其他特征,以及操作和运行结构的相关元件的方法与部件和制造经济的组合将变得更加显而易见,说明书中相同的附图标记表示各个附图中的相应部分。然而,应该清楚地理解,附图仅用于说明和描述的目的,并不旨在作为本发明的限制的定义。

### 附图说明

[0013] 本技术的各种实施例的某些特征在所附权利要求中特别提出。通过参考以下详细描述将获得对本技术的特征和优点的更好理解,所述详细描述提出了利用本发明的原理的示例性实施例,并且提供了附图:

[0014] 图1示出了根据各种实施例的关键短语识别的示例环境。

[0015] 图2示出了根据各种实施例的用于关键词语识别的示例系统。

[0016] 图3A-3B示出了根据各种实施例的用于关键词语识别的示例方法。

[0017] 图4A-4B示出了根据各种实施例的用于关键词语识别的示例方法的流程图。

[0018] 图5示出了示例计算机系统的框图,其中,可以实现本申请描述的任何实施例。

### 具体实施方式

[0019] 可以在各种情况下实现语音控制以便于用户控制。例如,通过安装在移动电话上的软件应用程序协调运输提供者(例如,驾驶员)和服务请求者(例如,乘客)的车辆服务平台可以通过将语音控制结合到应用程序来改进服务。在一个示例中,可以监视司机的讲话以确定他是否符合作业范围内的标准化语言要求。在另一个示例中,车辆乘员可以说预设的关键短语以触发机器检测要执行的命令。在紧急情况的一个示例中,车辆乘员可以通过说出某些SOS类型短语来寻求帮助,从而使机器检测并识别SOS短语并触发警报。在这样的示例中,语音识别是检测人类语音并将其转换为机器语言的基础,并且关键短语识别是机器识别与重要含义相关联的一部分流式语音的基础。为了增强语音识别的可靠性,期望提高正确接受率(即,在说出关键短语时,识别关键短语)并降低错误接受率(即,在未说出关键短语时,识别关键短语)。

[0020] 在各种实施方式中,计算系统可以被配置为识别关键短语,该关键短语由至少两个单词组成(例如,关键短语中以w1到w4的顺序的单词w1、w2、w3和w4)。当前技术将确定由四个候选词和概率p1、p2、p3和p4组成的候选短语,该四个概率分别指示候选词与关键词匹配的可能性,并且仅将候选短语的总概率 $\sqrt{p1p2p3p4}$ 与预设阈值进行比较以确定是否候选短语是关键短语。这种方法可能导致更高的错误接受率。例如,如果第一、第二和第四候选

词分别与w1、w2和w4高度匹配,即使第三候选词与w3不匹配且p3非常低,总概率也会超过预设阈值的总概率,将确定错误的关键词。作为另一示例,候选短语可以包括与关键字相同的单词,但是在不同的顺序次序中。由于总概率不考虑词序,所以候选短语的这种情况也会导致错误接受。

[0021] 基于计算机技术的要求保护的解决方案可以解决在关键词识别领域中出现的问题。除了确定关键词识别的总概率之外,计算机系统还可以测量各个候选词的概率。在四字关键词示例中,计算系统可以通过与对应的阈值进行比较来筛选每个候选词的个体概率(p1、p2、p3、p4)。如果p1、p2、p3和p4的任一概率均未能超过对应的阈值,则候选短语将不被确定为关键词,因此降低了错误接受率并增加了正确接受率。此外,计算系统可以考虑关键词识别的词序。计算机系统可以被配置用于计算接下来的单词的概率,其受先前单词的概率的影响。例如,p3的确定可以是有条件的或者可以直到确定p2超过其与关键词中的第二个词相关联的阈值才执行,从而消除诸如w1、w2、w4、w3之类短语的错误接受。

[0022] 本申请的各种实施例包括用于关键词识别的系统、方法和非暂时性计算机可读介质。用于关键词识别的示例性方法可以包括:获得包括音频部分的序列的音频;获得对应于至少两个音频部分的至少两个候选词,并获得所获得的候选词与音频部分之间的每个对应关系的第一概率分数;确定至少两个候选词是否分别匹配关键词的至少两个关键词,以及至少两个候选词的每个的第一概率分数是否超过对应的第一阈值,该至少两个候选词构成候选短语;响应于确定至少两个候选词匹配于至少两个关键词,以及每个第一概率分数超过对应的阈值,基于至少两个候选词的每个的第一概率分数,获取表示候选短语与关键词之间的匹配关系;并且响应于确定第二概率分数超过第二阈值,将候选短语确定为关键词。

[0023] 在一些实施例中,该方法可以由包括麦克风、处理器和存储指令的非暂时性计算机可读存储介质的移动设备实现。例如,麦克风可以被配置用于接收音频,并且指令由处理器执行时,使处理器执行该方法。所获得的音频可以包括由车辆中的一个或以上乘员的麦克风记录的语音。移动设备可以包括移动电话。

[0024] 在一些实施例中,获得对应于音频部分的至少两个候选词,并获得所获得的候选词与音频部分之间的每个对应关系的第一概率分数可包括,获得对应于音频的声谱图,沿着声谱图获得每个时间帧的特征向量,以获得对应于声谱图的至少两个特征向量,获得对应于至少两个特征向量的至少两个语言单元,至少基于将语言单元映射到单词的词典对应于音频的候选词序列,并且对于每个候选词,至少基于用语言单元的样本序列训练的模型获得第一概率分数,并从候选词序列中获得至少两个候选词。

[0025] 在一些实施例中,该方法可以进一步包括至少基于时间帧确定所获得的音频中的关键词的开始时间和结束时间。

[0026] 在一些实施例中,至少两个候选词可以是按时间顺序(即,从候选词序列和相同词序列中获得的连续词),并且至少两个候选词和至少两个关键词之间的分别匹配可以包括候选短语中的顺序次序中的候选词和关键词中的相同顺序次序中的关键词之间的匹配。

[0027] 在一些实施例中,确定至少两个候选词是否分别匹配关键词的至少两个关键词,以及至少两个候选词的每个的第一概率分数是否超过对应的第一阈值可以包括,在前向或后向顺序次序中确定至少两个候选词和至少两个关键词之间的分别匹配。

[0028] 在一些实施例中,该方法可以进一步包括:响应于确定至少两个候选词的任何第一个的第一概率分数不超过对应阈值,不将候选短语确定为关键短语。

[0029] 在一些实施例中,可以不基于语音模型或部分地基于语音模型来实现该方法,并且该方法可以不由语音解码器实现或部分地由语音解码器实现。

[0030] 在一些实施例中,该方法可以被实现为从音频中发现至少两个关键短语,并且至少两个关键短语可以包括用于唤醒应用程序的短语、标准化语言短语或紧急触发短语中的至少一个。实施例可以在各种场景中实现,例如在行走、打车,驾驶车辆、骑车等时,尤其是在打字不现实或不方便时。例如,为了监控车辆司机是否符合标准化服务语言,诸如移动电话或基于车辆的计算机之类的计算系统可以在与客户交谈期间监控司机的讲话。作为另一示例,用户可以对移动电话说出查询语句(例如,“XYZ, get me a ride to metro center”),使关键词短语“XYZ”唤醒应用程序并识别命令“get me a ride to metro center”。在另一个例子中,乘客的移动电话可以捕获请求帮助的语音(例如,“help”、“call 911”)并向适当方发送警报(例如,最近的警察巡逻车、警察局、医院、乘客的亲属)。因此,即使乘客卷入争斗、挣扎求生,或以其他方式无法拨打911或键入消息,也可以提醒适当方来进行救援。

[0031] 图1示出了根据各种实施例的用于关键词语识别的示例环境100。如图1所示,示例环境100可以包括至少一个计算系统102,其包括一个或以上处理器104和存储器106。存储器106可以是非暂时性的并且是计算机可读的。存储器106可以存储指令,当由一个或以上处理器104执行时,使得一个或以上处理器104执行本申请描述的各种操作。系统102还可以包括被配置用于捕获和记录音频输入(例如,人类语音或语音)的麦克风103。这里,可以使用任何其他替代音频捕获设备作为麦克风103。可以从计算设备107或用户101捕获音频输入。计算设备107(例如,手机、平板电脑、计算机、可穿戴设备(智能手表))可以向系统102发送和/或播放信息(例如,记录的音频)。用户101可以在麦克风103的检测范围内说话以进行音频捕获。可选地,系统102还可以包括显示器105,其被配置用于显示信息(例如,由系统102和/或计算设备107识别的语音文本)。显示器105可以包括触摸屏。系统102可以在诸如移动电话、平板电脑、计算机、可穿戴设备(智能手表)等各种设备上实现。

[0032] 上面的系统102可以安装有适当的软件(例如,应用程序、平台程序等)和/或硬件(例如,电线、无线连接等)以访问环境100的其他设备。环境100可以包括系统102可访问的一个或以上数据存储(例如,数据存储108)和一个或以上计算设备(例如,计算设备107)。在一些实施例中,系统102可以被配置为与数据存储108和/或计算设备107交换数据或信息。例如,数据存储108可以安装在用于存储地址信息的计算机中。计算设备107可以是被配置为执行语音识别的服务器。服务器可以被配置为从系统102接收音频输入并将各种模型(例如,隐马尔可夫模型(Hidden Markov Model)、基于动态时间扭曲的语音识别、神经网络)应用于音频输入以识别一个或以上的语音,并获取与语音相对应的文本。由于更高的计算能力,在计算设备107处执行的语音识别可以比在系统102处执行的语音识别更全面。对于本申请描述的应用,可以在系统102处执行关键短语识别,并且在一些情况下,响应于发现关键短语识别,可以将一个或以上的语音发送到计算设备107,以用于进一步的语音识别。

[0033] 在一些实施例中,数据存储108和/或计算设备107可以实现在线信息或服务平台。该服务可以与车辆(例如,汽车、自行车、船、飞机等)相关联,并且该平台可以被称为车辆

(服务)呼叫平台。平台可以接受运输请求、识别满足请求的车辆、安排接驾和处理交易。例如,用户可以使用系统102(例如,安装有与平台相关联的应用程序的移动电话)来向平台提交运输请求。计算设备107可以接收和发布运输请求。车辆司机可以使用系统102(例如,安装有与平台相关联的应用程序的另一个移动电话)来接收所发布的交通请求并获得接驾位置信息和用户的信息。一些平台数据(例如,车辆信息、车辆司机信息、地址信息等)可以存储在存储器106中,或者可以从数据存储108和/或计算设备107中检索。如本申请所述,用于关键短语识别的系统102可以与人或车辆相关联(例如,由司机携带、由乘客携带、由与车辆无关的人使用,安装在车辆中)或以其他方式能够捕获访问平台的人的语音。

[0034] 在一些实施例中,系统102和一个或以上计算设备(例如,计算设备107)可以集成在单个设备或系统中。或者,系统102和计算设备可以作为单独的设备操作。数据存储可以是系统102可访问的任何地方,例如,在存储器106中,在计算设备107中,在耦合到系统102的另一设备(例如,网络存储设备)中,或另一个存储位置(例如,基于云的存储系统、网络文件系统等)等。尽管在该图中系统102和计算设备107被示为单个组件,但是应当理解,系统102和计算设备107可以实现为耦合在一起的单个设备或多个设备。计算设备可以耦合到诸如系统102的多个系统并与之交互。通常,系统102、计算设备107和数据存储108能够通过一个或以上可以传送数据的无线网络(例如,因特网)互相通信。下面参考图2至图5描述环境100的各个方面。

[0035] 图2示出了根据各种实施例的用于关键词语识别的示例系统200。图2中所示并在下面呈现的操作是说明性的。除了为简单起见而移除数据存储108和计算设备107,图2中的各种设备和组件与图1中描述的那些类似。

[0036] 在各种实施例中,系统102可以在包括移动电话的移动设备上实现。系统102的一个或以上组件(例如,麦克风108、处理器104和/或存储器106)可以被配置用于接收音频(例如,音频202)并将音频存储在音序列(例如,作为存储器106中的文件)。当音频被连续流传输时,这里描述的关键短语识别方法可以对音序列执行,或者可以在流传输完成之后对音序列执行。音频202可以包括人类所说的语音(例如,句子、短语、单词)。语音可以是任何语言。处理器104可以被配置用于控制记录的开始和停止。例如,当如上所述进入安装在系统102上的应用程序的预设界面或打开应用程序时,可以开始记录。处理器104可以控制系统102的模数转换器(ADC)(在该图中未示出)以将捕获的音频转换为数字格式并存储在音序列中。音序列可以与时间相关联,并且可以包括捕获的音频的时间序列数据。音序列可以以各种音频文件格式(例如,WAV文件)存储。音序列可以存储在存储器106中、高速缓存中或其他存储介质中。音序列可以不限于特定的操作系统,并且可以使用各种替代音频缓冲器、音频高速缓存、音频流或音频回调技术来代替音序列。音序列可以可选地被配置为仅捕获最新音频(例如,最后一分钟捕获的音频、最后1G音频文件、一天中捕获的音频)。例如,捕获的音频可以连续地流传输到有限大小的高速缓存,并且超过限制的最近音频部分被写在最老的音频部分上。

[0037] 在一些实施例中,系统102的一个或以上组件(例如,处理器104和/或存储器106)可以被配置用于监视音序列,以发现一个或以上关键短语。响应于发现关键短语,处理器104和/或存储器106可被配置为可选地将信息204发送到计算设备107。例如,响应于系统102发现唤醒短语,可以触发系统102上的应用的功能以从音序列获得音频片段,并且应

用可以将获得的音频片段作为信息204发送到计算设备107用于语音识别。作为另一个示例,系统102可以捕获持续时间的语音(例如,一天中司机的讲话)并将语音作为信息204发送到计算设备107以用于语音识别。语音识别可用于确定司机所捕获的语音符合标准的程度。作为又一示例,系统102可以实时(例如,每几毫秒)监视捕获的音频,发现关键短语(例如,标准化短语、紧急求助音素),并发送信息204(例如,符合语言标准的指示、紧急警报)到计算设备107。计算设备107可以与适当方相关联,例如司机性能评估器、客户服务人员、救援人员、警察等。

[0038] 在一些实施例中,计算设备107可以将信息206(例如,由计算设备107识别的语音的文本)返回给系统102。系统102的显示器105可以被配置用于显示返回的信息。

[0039] 图3A和3B示出了根据各种实施例的用于关键词语识别的示例方法300。方法300可以在各种环境中实现,包括例如图1的环境100。示例方法300可以由系统102的一个或以上组件(例如,处理器104、存储器106)实现。以下呈现的方法300的操作是说明性的。取决于实现方式,示例方法300可以包括以各种顺序或并行执行的附加的、更少的或替代的步骤。

[0040] 在一些实施例中,音频序列301-303可以表示由系统102捕获的示例音频。音频序列在x轴方向上以连续时间序列用对应的语音字块,暂停(pau)或静音(sil)标记。如下面从步骤305所示,这些标签将从捕获的音频中确定。一些语音词可以是关键短语的关键词。例如,在音频序列301中,“call the police”是被识别以触发紧急警报的关键短语。在音频序列302中,“how can I help you”是要检测的标准化语言,其可用于对司机提供的服务进行评分。在音频序列303中,“hello my device”是用于触发应用程序、进程或功能的唤醒短语。在检测到该唤醒短语时,系统102可以捕获下一个句子(例如,“order coffee to my car”)。系统102可以识别该句子,并基于该查询执行(例如,通过向商家下达咖啡的递送订单,并提供汽车的位置)。或者,系统102可以将句子的音频发送到计算设备107,使得计算设备107识别该句子并基于该查询执行。与其他关键短语一样,唤醒短语可以包括一个或以上的单词。唤醒短语可以包括名称或问候语(例如,“Hello XYZ”,“Ok XYZ”,“XYZ”,“Hello my device”)并且可以与应用程序、程序、功能、过程或设备(例如,应用程序XYZ、我的设备)。这里,“唤醒”不一定意味着从“睡眠模式”唤醒。在“唤醒”之前,系统102可以休眠、空闲或执行其他任务。

[0041] 以下参考图3A-3B的示例性步骤和说明主要与示例音频序列303一致。在一些实施例中,关键短语可以由至少两个关键词组成,每个关键词与第一阈值相关联。对于每个关键字,阈值可以设置用于确定从音频中发现的候选词是关键词的最小概率。例如,“hello”,“my”和“device”中的每个可能需要至少80%的概率来确定候选词的匹配。下面描述候选词的确定。以下各种实施例可以使用“hello my device”作为关键短语,并且通过所公开的方法,可以相应地获得“hello my device”的“一个候选短语”或“至少两个候选词”。“至少两个候选词”可以包含在“候选词序列”中。

[0042] 通过将单词分解为语言单元,音频序列304是音频序列303的替代表示。语言单元可以有許多分类和定义,例如音素、音素部分、三音素、单词、n元语法等。在一个示例中,音素是具有语言中的独特含义或功能的语音组,并且可以是语言的音韵中最小的有意义的对比单元。音素的数量因语言而异,大多数语言具有20-40个音素。美式英语可具有大约40个音素(24个辅音,16个元音)。在一个示例中,“hello”可以被分成语言单元/音素“hh”、“ah”、

“l”和“ow”。类似地,音频序列303可以由语言单元队列304表示。

[0043] 在步骤305,可以使用频谱来表示所获得的音频(例如,音频202)。音频可以有各种不同的表示。在该示例中,频谱可以示出捕获的声音相对于时间的幅度。在各种实施例和实现中,由处理器获得的音频可以对应于频谱或替代格式,并且将如下关于步骤306-311所讨论的进行处理。通过利用这些步骤处理频谱,可以获得语言单元队列304中的对应语言单元和音频序列303中的对应字序列。图3A中的垂直虚线可以在各种图示上标记相同的时间戳,并指示它们之间的对应关系。步骤306-311也可以被称为示例性声模型,具有来自传统方法的修改和改进。

[0044] 在步骤306,可以至少基于频谱获得声谱图。声谱图可以是语音信号的频率对时间的表示。在一些实施例中,傅里叶变换可以应用于来自步骤305的声谱以获得声谱图。在声谱图中,幅度信息以灰度显示为暗区和亮区。亮区域可以指示在对应频率的对应时间没有捕获声音(例如,暂停、静音),暗区域可以指示存在声音。基于x轴方向上的暗图案和亮图案的变化,可以确定语言单元(例如,单词、音素)之间的边界。因此,在确定关键短语之后,可以相应地确定关键短语的开始和结束时间戳。此外,两条虚线之间的y轴方向上的暗区域的图案可以指示在对应时间段捕获的各种频率,并且可以提供共振峰(携带该声音的身份)和过渡的信息以帮助确定对应的手机。

[0045] 在步骤307,可以至少基于声谱图获得各种特征向量。在一些实施例中,倒谱分析可以应用于声谱图以获得特征向量。例如,时间帧(例如,25毫秒长的时间帧)可以沿着时间轴移动并且通过帧(每10毫秒一个帧)对声谱图帧进行采样。本领域技术人员将理解诸如梅尔频率倒谱系数(MFCC)的技术的应用(即,将倒谱分析应用梅尔频谱以获得多个梅尔频率倒谱系数(MFCC))。因此,在步骤307,语音信号可以被转换为显示为矩形块的一系列矢量。每个矢量可以是39D矢量,包括12个梅尔频率倒谱系数(MFCC)特征、12个增量梅尔频率倒谱系数(MFCC)特征、12个增量-增量梅尔频率倒谱系数(MFCC)特征、1个(log)帧能量、1个增量(log)帧能量和1个增量-增量(log帧能量)。可以将这些向量给予模式分类器以识别对应的语言单元。

[0046] 在步骤308,可以基于各种模型方法将特征向量映射到各种语言单元,例如基于码本的模式分类,取决于上下文的三音素模型等。通过用语言的样本记录训练模型,该模型可用于将输入特征向量分类为语言单元之一。因此,可以相应地获得在步骤308以小块示出的候选音素(或其他语言单元),诸如“hh”和“ah”。

[0047] 继续图3B的方法300,在步骤309,可以使用诸如隐马尔可夫模型(HMM)的各种方法来至少基于所获得的语言单元获得候选词。对于要发现的有限数量的关键短语(例如,“Hello my device”,“call the police”,“how can I help you”),可以为每个关键词建立一个HMM关键短语。词典可用于将每个关键词映射到一个或以上语言单元。例如,“low”可以映射到音素“l”和“ow”,并且“hello”可以映射到音素“hh”、“ah”、“l”和“ow”。如图所示。在步骤309的表格中,x轴表示以对应于音频的时间序列从特征向量获得的一系列语言单元,并且y轴表示具有对应语言单元的各种单词(例如,关键词)。诸如“A1”和“B3”的标签表示从特征向量获得的语言单元与关键词的语言单元之间的对应关系。对于获得的每个语言单元,它可以是几个不同单词的一部分(例如,“ow”可以是“low”或“hello”的一部分)。语音的样本记录训练的HMM可以通过分配从一个语言单元到另一个语言单元的转换之间的概率

来模拟语音。也就是说,链接候选语言单元的箭头可以各自携带概率,类似于决策树。箭头可以是向前或向后的。这里,从A4到A3、从A3到A2、以及从A2到A1的概率和这些概率的乘积可以高于其他选择(例如,B4、B3、B2、B1之间的链接),呈现“hello“作为对应于”hh-ah-l-ow“的最高概率的词。由于不常见的使用,其他选择的概率可能较低,可以假设反映在训练样本中。因此,在步骤309,可以获得对应于音频部分的序列的候选词序列(每个音频部分可以对应于音频中语言单元的一个或以上)(例如,“hello my device pau order coffee to my car pau sil”),并且可以获得所获得的候选词和音频部分之间的每个对应关系的第一概率分数。因此,根据候选词的序列,可以获得对应于音频部分的至少两个候选词,其中音频部分的序列包括至少两个音频部分。

[0048] 在步骤310,如果从候选词序列获得的至少两个候选词(该至少两个候选词构成候选短语)与关键短语中的关键词序列匹配,则可以确定第一概率分数(置信度)以确定该匹配。例如,观察诸如 $P(\text{hello})$ 、 $P(\text{my})$ 和 $P(\text{device})$ 的候选词的概率可以分别与关键词“hello”、“my”和“device”相关的第一阈值进行比较。观察候选词的概率可以以各种方式确定。例如, $P(\text{hello})$ 可以被确定为上述步骤309中的 $P(A4\text{至}A3)$ 、 $P(A3\text{至}A2)$ 和 $P(A2\text{至}A1)$ 的乘积。如果候选词概率中的任何一个没有超过对应的第一阈值,则可以不将至少两个候选词(即,候选短语)确定为关键短语。因此,即使确定下面参考步骤311描述的“hello me device”被满足,诸如“hello me device”的短语也可以被适当地确定为与关键短语不匹配。

[0049] 此外,候选词的第一概率分数可以根据它们在候选短语中的顺序(例如,以前向或后向顺序)来顺序地测量。例如,在正向顺序中, $P(\text{device})$ 可以不与其阈值进行比较,直到确定 $P(\text{my})$ 超过其阈值。在另一示例中,候选短语中的第N个候选词仅与关键短语中的第N个关键词的阈值进行比较。因此,即使下面参考步骤311描述的确定“hello device my”被满足,诸如“hello device my”的短语也可以被正确地确定为与关键短语不匹配。这里,候选短语可能不一定是人们通常使用的短语。它可以包括预设的组成短语或公共短语。

[0050] 在步骤311,可以将候选短语的总概率与与关键短语相关联的第二阈值进行比较。例如,可以将 $P(\text{hello})P(\text{my})P(\text{device})$ 的平方根与第二阈值0.5进行比较。如果超过第二阈值,则可以将候选短语确定为关键短语。即使执行步骤310以确保短语的整体匹配,该步骤也是必要的,尤其是当步骤310中的第一阈值相对较低时。如果未超过第二阈值,则可以不将候选短语确定为关键短语。

[0051] 如上所述,用于关键词语识别的示例性方法可以包括:获得包括音频部分的序列(例如,上述步骤305,该部分可以是该音频的任何部分,例如对应于单词的一段)的音频(例如,上述音频202);获得对应于至少两个音频部分的至少两个候选词,并获得所获得的候选词与音频部分之间的每个对应关系的第一概率分数(例如,上述步骤306-309,其中从一个音频部分的序列中获得一系列候选词,其中包括至少两个候选词的候选词序列和包含至少两个音频部分的音频部分的序列);确定至少两个候选词是否分别匹配关键短语的至少两个关键词,以及至少两个候选词的每个的第一概率分数是否超过对应的第一阈值(例如,步骤如上所述的310,至少两个候选词构成候选短语(例如,上面描述的“hello my device”);响应于确定至少两个关键词匹配至少两个候选词且每个第一概率分数超过对应阈值,基于至少两个候选词的每个的第一概率分数,获得表示该候选短语和关键短语之间的匹配关系

的第二概率分数;并且响应于确定第二概率得分超过第二阈值,将候选短语确定为关键短语(例如,上述步骤311)。该方法可以进一步包括:响应于确定不超过对应阈值的任何至少两个候选词的第一概率分数,不将候选短语确定为关键短语。

[0052] 在一些实施例中,该方法可以由包括麦克风、处理器和存储指令的非暂时性计算机可读存储介质的移动设备实现。例如,麦克风可以被配置用于接收音频,并且当由处理器执行时,指令可以使处理器执行该方法。获得的音频可以包括由麦克风记录的语音。可以从车辆中的一个或以上乘员(例如,司机、乘客)捕获语音。移动设备可以包括移动电话。

[0053] 在一些实施例中,音频部分的序列可包括时间序列。获取对应于至少两个音频部分的至少两个候选词,并获得所获得的候选词与音频部分之间的每个对应关系的第一概率分数,可以包括:获得对应于该音频(例如,上述步骤306)的声谱图,沿着声谱图获得每个时间帧的特征向量,以获得对应于声谱图(例如,上述步骤307)的至少两个特征向量,获得对应于至少两个特征向量(例如,上述步骤308)的至少两个语言单元,至少基于将语言单元映射到单词的词典,获得对应于音频的候选词序列,以及对于每个候选者单词,至少基于用语言单元的样本序列训练的模型获得第一概率分数(例如,上述步骤309);并且从候选词的序列获得至少两个候选词(例如,在上述步骤309,可以获得包括至少两个候选词的候选短语作为候选词序列的一部分)。

[0054] 在一些实施例中,该方法可以进一步包括至少基于时间帧确定所获得的音频中的关键短语的开始时间和结束时间。例如,上述步骤306中描述的模式边界处可用于确定各种语言单元的开始和结束时间。对于参考上述步骤306至308的另一示例,在每个时间帧,可以基于诸如应用由样本语言训练的模型的方法确定所获得的特征向量和语言单元之间的匹配的概率分数。较高分数可以指示语言单元的开始或结束,其可以链接到所获得的候选词,以帮助确定音频中的候选短语的开始和结束时间。

[0055] 在一些实施例中,至少两个候选词可以是按时间顺序(即,从候选词序列和相同词序列中获得的连续词),以及至少两个候选词与至少两个关键词之间的分别匹配可以包括候选短语中的顺序次序中的候选词与关键短语中的相同顺序次序中的关键词之间的匹配。例如,获得的音频是“what a nice day today,hello my device,order coffee to my car...”,其中关键短语是“hello my device”。从该获得的音频,可以获得一个或以上候选词例如“hat a nice day today,hello my device,order coffee to my car...”,并且在步骤309可以从中获得一个候选短语“hello my device”。为了确保候选词和关键词(相同数量的词)之间的匹配是准确的,可以顺序地比较候选短语中的候选词和关键词短语中的关键词。也就是说,可以将候选短语中的第一个单词与关键短语中的第一个单词进行比较,可以将候选短语中的第二个单词与关键短语中的第二个单词进行比较,等等,直到所有的单词都被比较。如果确认候选短语和关键短语包括相同序列中的相同词,则可以评估第一和第二概率得分以进一步确定匹配。

[0056] 在一些实施例中,确定至少两个候选词是否分别匹配关键短语的至少两个关键词,以及至少两个候选词的每个的第一概率分数是否超过对应的第一阈值,则可以包括在前向或后向顺序次序(在上述步骤310中显示为向前和向后箭头)中确定至少两个候选词和至少两个关键词之间的分别匹配。

[0057] 在一些实施例中,可以不基于语音模型或部分地基于语音模型来实现该方法,并

且该方法可以不由语音解码器实现或部分地由语音解码器实现。如上所述,由于需要发现固定数量的关键短语,因此这些关键短语可以是单独的模型并且基于声学来确定,从而避免使用语音模型和语音解码器。在没有显著的计算能力要求的情况下,所公开的方法和系统可以在诸如移动电话的移动设备上实现。移动设备可以在不从诸如服务器的其他设备获取计算能力的情况下发现关键短语。在一些实施例中,该方法可以被实现为从音频中发现至少两个关键短语,并且至少两个关键短语可以包括用于唤醒应用的短语、标准化语言的短语或者紧急触发短语中的至少一个。

[0058] 图4A示出了根据本申请的各种实施例的用于关键词语识别的示例方法400的流程图。方法400可以在各种环境中实现,包括例如图1的环境100。示例方法400可以由系统102的一个或以上组件(例如,处理器104、存储器106)实现。以下呈现的方法400的操作是说明性的。取决于实现方式,示例方法400可以包括以各种顺序或并行执行的附加、更少或替代步骤。方法400可以包括以下步骤(步骤被示为框)。

[0059] 在框401处,可以获得包括音频部分的序列的音频。在框402,可以获得对应于至少两个音频部分的至少两个候选词,并且可以获得所获得的候选词和音频部分之间的每个对应关系的第一概率分数。在框403,可以确定至少两个候选词是否分别匹配关键短语的至少两个关键词,以及至少两个候选词的每个的第一概率分数是否超过对应的第一词,至少两个候选词构成候选短语。在框404,响应于确定至少两个候选词匹配至少两个关键词且每个第一概率分数超过对应的阈值,基于至少两个候选词的每个的第一概率分数,可以获得表示候选短语与关键短语之间的匹配关系的第二概率分数。在框405,响应于确定第二概率得分超过第二阈值,可以将候选短语确定为关键短语。

[0060] 图4B示出了根据本申请的各种实施例的用于关键短语识别的示例方法410的流程图。方法410可以在各种环境中实现,包括例如图1的环境100。示例方法410可以由系统102的一个或以上组件(例如,处理器104、存储器106)实现。以下呈现的方法410的操作是说明性的。取决于实现方式,示例方法410可以包括以各种顺序或并行执行的附加、更少或替代步骤。上述块402可以包括方法410。方法410可包括以下步骤。

[0061] 在框411,可以获得对应于音频(在上面的框401中描述)的声谱图。在框412处,可以沿着声谱图为每个时间帧获得特征向量,以获得对应于声谱图的至少两个特征向量。在框413,可以获得对应于至少两个特征向量的至少两个语言单元。在框414,可以至少基于将语言单元映射到单词的词典获得对应于音频的候选词序列,并且对于每个候选词,至少基于由语言单元的样本序列训练的模型,获得第一概率分数。在框415,可以从候选词序列获得至少两个候选词。

[0062] 这里描述的技术由一个或以上专用计算设备实现。专用计算设备可以包括一个或以上硬件处理器,其被编程为根据固件、存储器、其他存储器或组合中的程序指令来执行技术。专用计算设备可以是台式计算机系统、服务器计算机系统、便携式计算机系统、手持设备、网络设备或包含硬连线和/或程序逻辑以实现该技术的任何其他设备或设备的组合。计算设备通常由操作系统软件控制和协调。传统操作系统控制和调度用于执行的计算机进程、执行存储器管理、提供文件系统、网络、输入/输出服务、以及提供用户界面功能,例如图形用户界面(Graphical User Interface,“GUI”)等的东西。

[0063] 图5是示出计算机系统500的框图,在该计算机系统500上可以实现本申请描述的

任何实施例。系统500可以对应于上述系统102。计算机系统500包括总线502或用于传送信息的其他通信机制,一个或以上硬件处理器504与总线502耦合以处理信息。硬件处理器504可以是例如一个或以上通用微处理器。处理器504可以对应于上述处理器104。

[0064] 计算机系统500还包括主存储器506,例如随机存取存储器(RAM)、高速缓存和/或其他动态存储设备,耦合到总线502,用于存储将由处理器504执行的信息和指令。主存储器506还可以用于在执行要由处理器504执行的指令期间存储临时变量或其他中间信息。当存储在处理器504可访问的存储介质中时,这样的指令使计算机系统500进入专用机器,该专用机器被定制以执行指令中指定的操作。计算机系统500还包括只读存储器(ROM)508或耦合到总线502的其他静态存储设备,用于存储处理器504的静态信息和指令。提供存储器510,例如磁盘、光盘或USB拇指驱动器(闪存驱动器)等,并将其耦合到总线502,用于存储信息和指令。主存储器506、只读存储器(ROM)508和/或存储器510可以对应于上述存储器106。

[0065] 计算机系统500可以使用定制的硬连线逻辑、一个或以上专用集成电路或现场可编程门阵列、固件和/或程序逻辑来实现本申请描述的技术,其与计算机系统组合使计算机系统500成为-专用机器。根据一个实施例,本申请的技术由计算机系统500执行,以响应处理器504执行包含在主存储器506中的一个或以上指令的一个或以上序列。这些指令可以从另一存储介质(例如存储器510)读入主存储器506。包含在主存储器506中的指令序列的执行使得处理器504执行本申请描述的处理步骤。在替代实施例中,可以使用硬连线电路来代替软件指令或与软件指令组合。

[0066] 主存储器506、只读存储器508和/或存储器510可以包括非暂时性存储介质。这里使用的术语“非暂时性介质”和类似术语是指存储使机器以特定方式操作的数据和/或指令的任何介质。这种非暂时性介质可包括非易失性介质和/或挥发性介质。非易失性介质包括例如光盘或磁盘,例如存储器510。易失性介质包括动态存储器,例如主存储器506。非暂时性介质的常见形式包括,例如,软磁盘、软盘、硬盘、固态驱动器、磁带、或任何其他磁数据存储介质、只读光盘驱动器、任何其他光学数据存储介质、具有孔图案的任何物理介质、随机存取存储器、可编程序只读存储器和可擦编程只读存储器、快可擦编程只读存储器、非易失随机存取存储器、任何其他存储器芯片或盒式磁带,以及它们的网络版本。

[0067] 计算机系统500还包括麦克风512或替代音频捕获设备。麦克风512可以对应于上述麦克风103。

[0068] 计算机系统500还包括耦合到总线502的通信接口518。通信接口518提供耦合到连接到一个或以上本地网络的一个或以上网络链路的双向数据通信。例如,通信接口518可以是综合业务数字网(ISDN)卡、电缆调制解调器、卫星调制解调器或调制解调器,以提供与对应类型的电话线的数据通信连接。又例如,通信接口518可以是局部区域网络(LAN)卡,以提供到兼容局部区域网络(LAN)(或与广域网(WArea Network)通信的广域网(Wide Area Network))的数据通信连接。还可以实现无线链路。在任何这样的实现中,通信接口518发送和接收携带表示各种类型信息的数字数据流的电信号、电磁信号或光信号。

[0069] 计算机系统500可以通过网络、网络链路和通信接口518发送消息并接收包括程序代码的数据。在因特网示例中,服务器可以通过因特网、因特网服务提供者(ISP)、本地网络和通信接口518发送用于应用程序的所请求的代码。

[0070] 接收到的代码可以在处理器504被接收时执行,和/或存储在存储器510或其他非

易失性存储器中,以供稍后执行。

[0071] 前面部分中描述的每个过程、方法和算法可以由一个或以上计算机系统或包括计算机硬件的计算机处理器执行的代码模块实现,并且完全或部分自动化。可以部分地或全部地在特定于应用的电路中实现过程和算法。

[0072] 上述各种特征和过程可以彼此独立地使用,或者可以以各种方式组合。所有可能的组合和子组合旨在落入本公开的范围。另外,在一些实现中可以省略某些方法或过程框。本申请描述的方法和过程也不限于任何特定序列,并且与其相关的块或状态可以以适当的其他序列执行。例如,所描述的块或状态可以以不同于具体公开的顺序执行,或者多个块或状态可以在单个块或状态中组合。示例性块或状态可以串行、并行或以某种其他方式执行。可以将块或状态添加到所公开的示例性实施例或从其中移除。这里描述的示例系统和组件可以与所描述的不同地配置。例如,与所公开的示例性实施例相比,可以添加、移除或重新布置元件。

[0073] 本申请描述的示例方法的各种操作可以至少部分地由临时配置(例如,通过软件)或永久地配置用于执行相关操作的一个或以上处理器来执行。无论是临时配置还是永久配置,这样的处理器可以构成处理器实现的引擎,其操作以执行本申请描述的一个或以上操作或功能。

[0074] 类似地,这里描述的方法可以至少部分地由处理器实现,其中特定处理器是硬件的示例。例如,方法的至少一些操作可以由一个或以上处理器或处理器实现的引擎来执行。此外,一个或以上处理器还可以操作以支持“云计算”环境中的相关操作的性能或“软件即服务”(SaaS)。例如,至少一些操作可以由一组计算机(作为包括处理器的机器的示例)执行,这些操作可以经由网络(例如,因特网)和通过一个或以上适当的接口(例如,应用程序接口(API))来访问。

[0075] 某些操作的性能可以在处理器之间分配,不仅驻留在单个机器内,而且部署在多个机器上。在一些示例实施例中,处理器或处理器实现的引擎可以位于单个地理位置(例如,在家庭环境、办公室环境或服务器群内)。在其他示例实施例中,处理器或处理器实现的引擎可以分布在多个地理位置。

[0076] 贯穿本说明书,多个实例可以实现被描述为单个实例的组件、操作或结构。尽管一个或以上方法的各个操作被示出并描述为单独的操作,但是可以同时执行各个操作的一个或以上,并且不要求以所示的顺序执行操作。作为示例配置中的单独组件呈现的结构和功能可以实现为组合结构或组件。类似地,作为单个组件呈现的结构和功能可以实现为单独的组件。这些和其他变化、修改、添加和改进落入本申请主题的范围。

[0077] 尽管已经参考特定示例实施例描述了主题的概述,但是在不脱离本申请的实施例的更宽范围的情况下,可以对这些实施例进行各种修改和改变。本申请中的主题的这些实施方案可以单独地或共同地通过术语“发明”在本申请中仅为了方便而被引用,并且不意图将本申请的范围自愿地限制于任何单个公开或概念,如果实际不止一个已披露。

[0078] 本申请所示的实施例得到足够详细的描述,以使本领域技术人员能够实践所公开的指导。可以使用其他实施例并从中导出其他实施例,使得可以在不脱离本公开的范围的情况下进行结构和逻辑替换和改变。因此,具体实施方式不应被视为具有限制意义,并且各种实施例的范围仅由所附权利要求以及这些权利要求所赋予的等价物的全部范围来限定。

[0079] 在此描述和/或在附图中描绘的流程图中的任何过程描述、元素或块应该被理解为潜在地表示包括用于实现该过程中特定逻辑或步骤的一个或以上可执行指令的模块、片段或代码部分。替代实施方式包括在本申请描述的实施方案的范围内,其中可以从所示或所讨论的顺序中删除、执行元件或功能,包括基本上同时或以相反的顺序,取决于所涉及的功能,如本领域技术人员可以理解。

[0080] 如本申请所用,术语“或”可以以包含性或排他性的含义来解释。此外,可以为在此描述为单个实例的资源、操作或结构提供多个实例。另外,各种资源、操作、引擎和数据存储之间的边界在某种程度上是任意的,并且在特定说明性配置的上下文中示出了特定操作。设想了其他功能分配,并且可以落入本申请的各种实施例的范围内。通常,在示例配置中作为单独资源呈现的结构和功能可以实现为组合结构或资源。类似地,作为单个资源呈现的结构和功能可以实现为单独的资源。这些和其他变化、修改、添加和改进落入由所附权利要求表示的本申请的实施例的范围内。因此,说明书和附图应被视为说明性的而非限制性的。

[0081] 条件语言,例如,其中,“可以”、或“可以”,除非另有明确说明,或者在所使用的上下文中以其他方式理解,通常旨在表达某些实施例包含某些特征、元件和/或步骤,而其它实施例不包括。因此,这种条件语言通常不旨在暗示一个或以上实施例以任何方式需要特征、元素和/或步骤,或者一个或以上实施例必须包括用于决定的逻辑,无论是否有用户输入或提示,是否包括这些特征、元素和/或步骤,或者是否在任何特定实施例中执行。

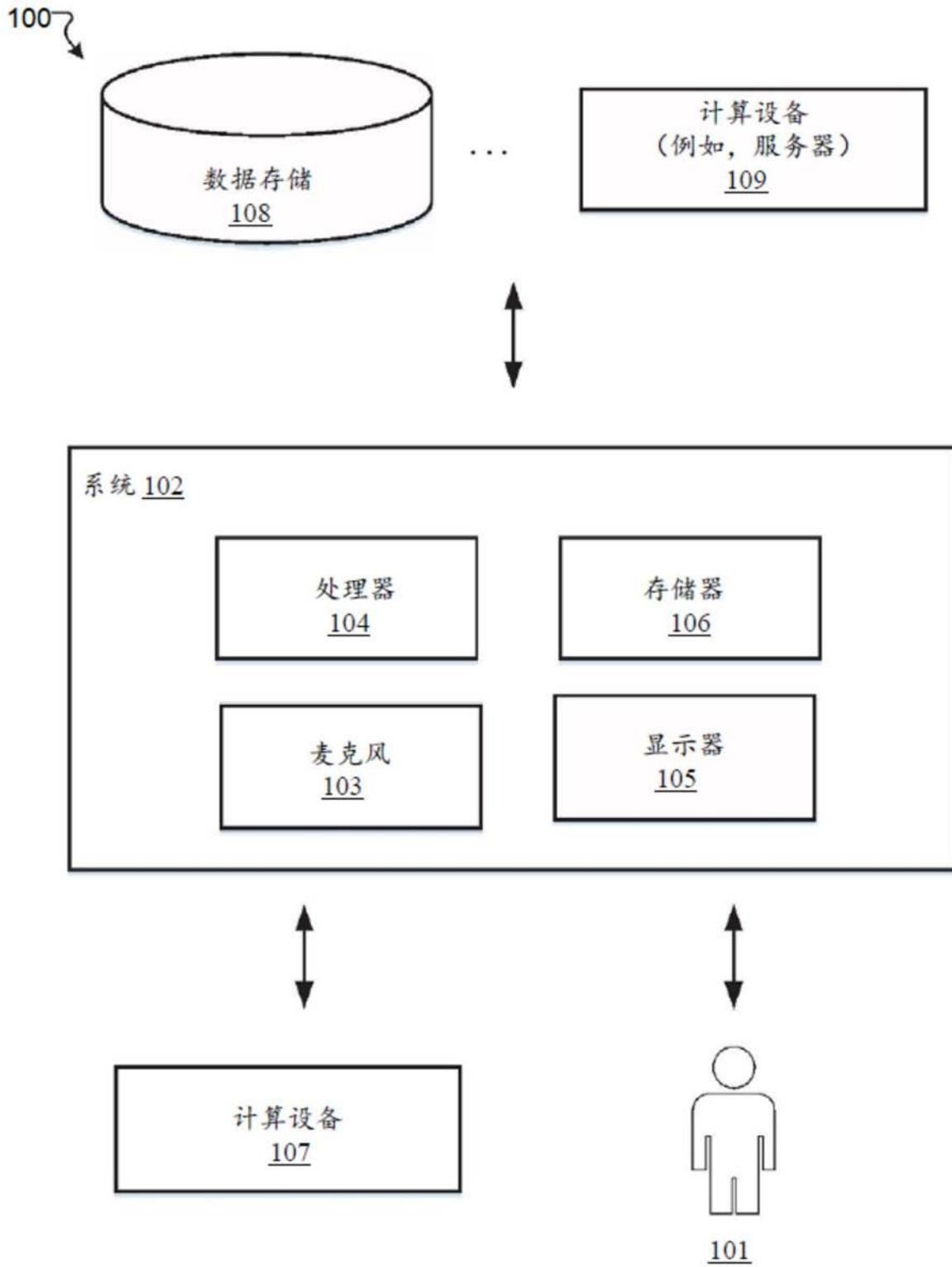


图1

200 ↘

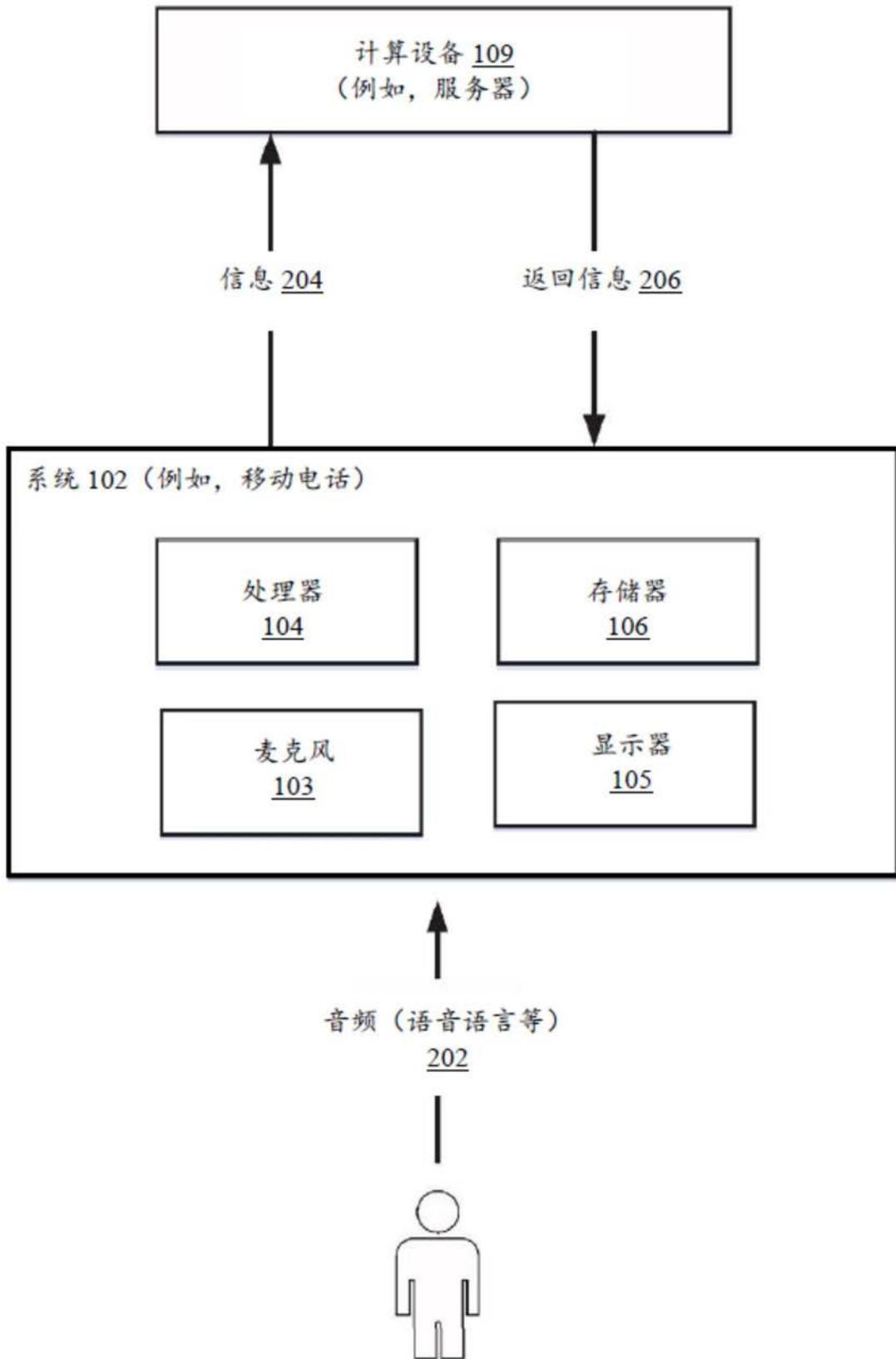


图2

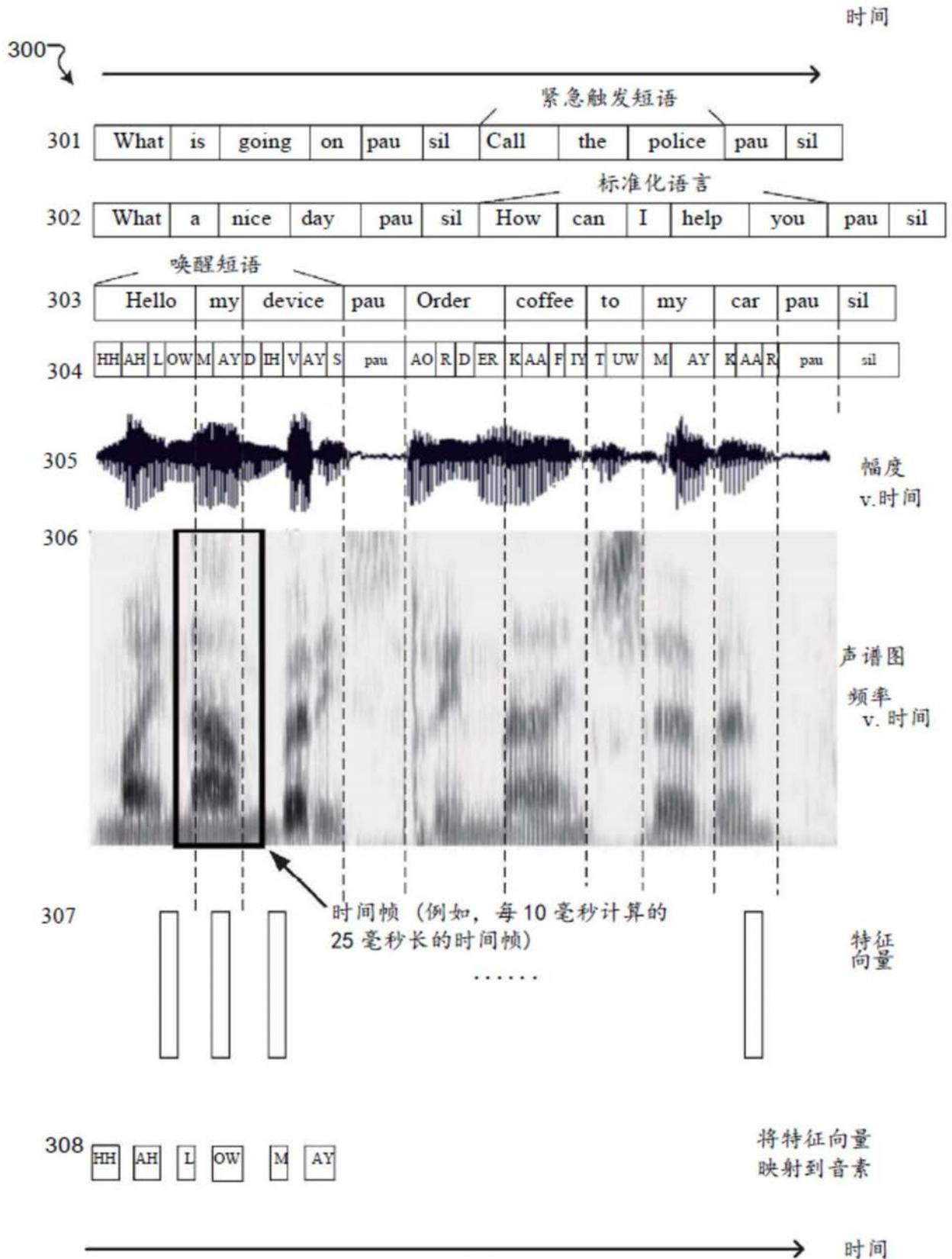
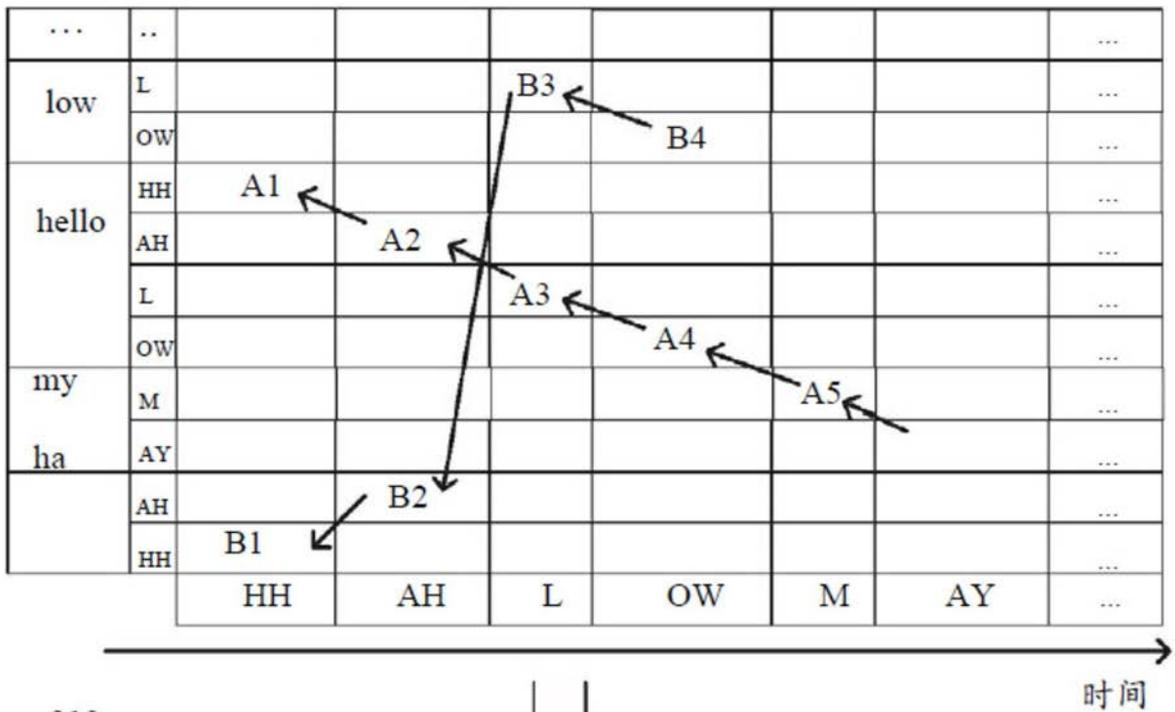


图3A

309

续图 3A



310

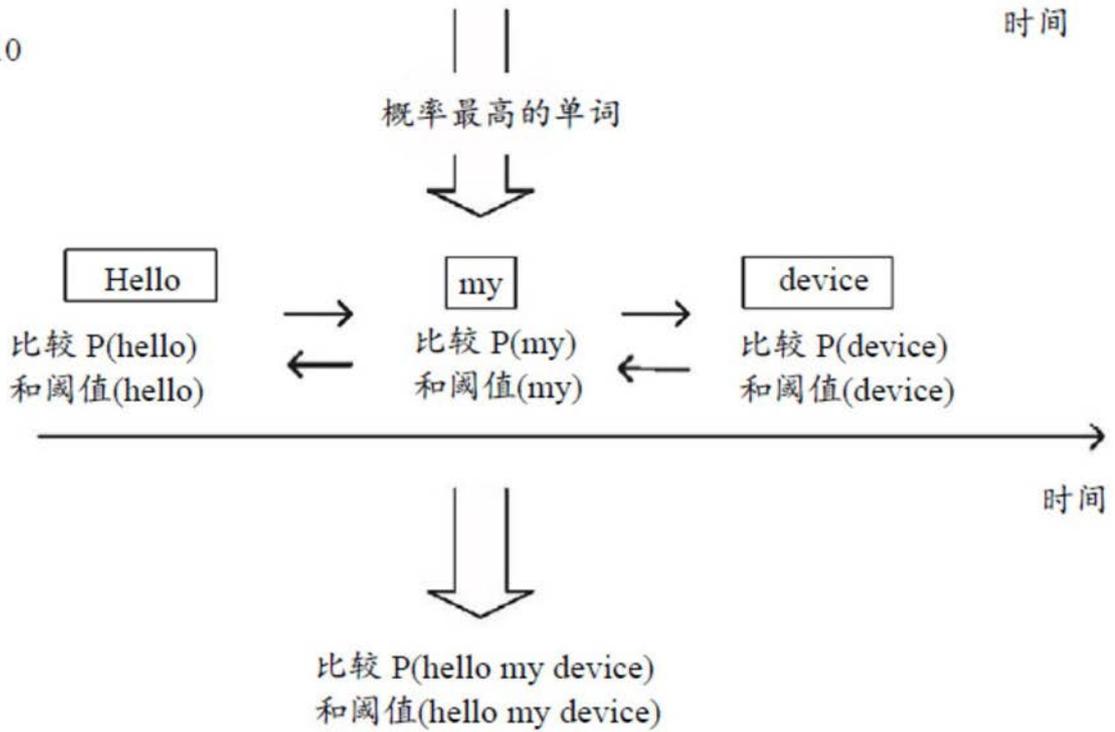


图3B

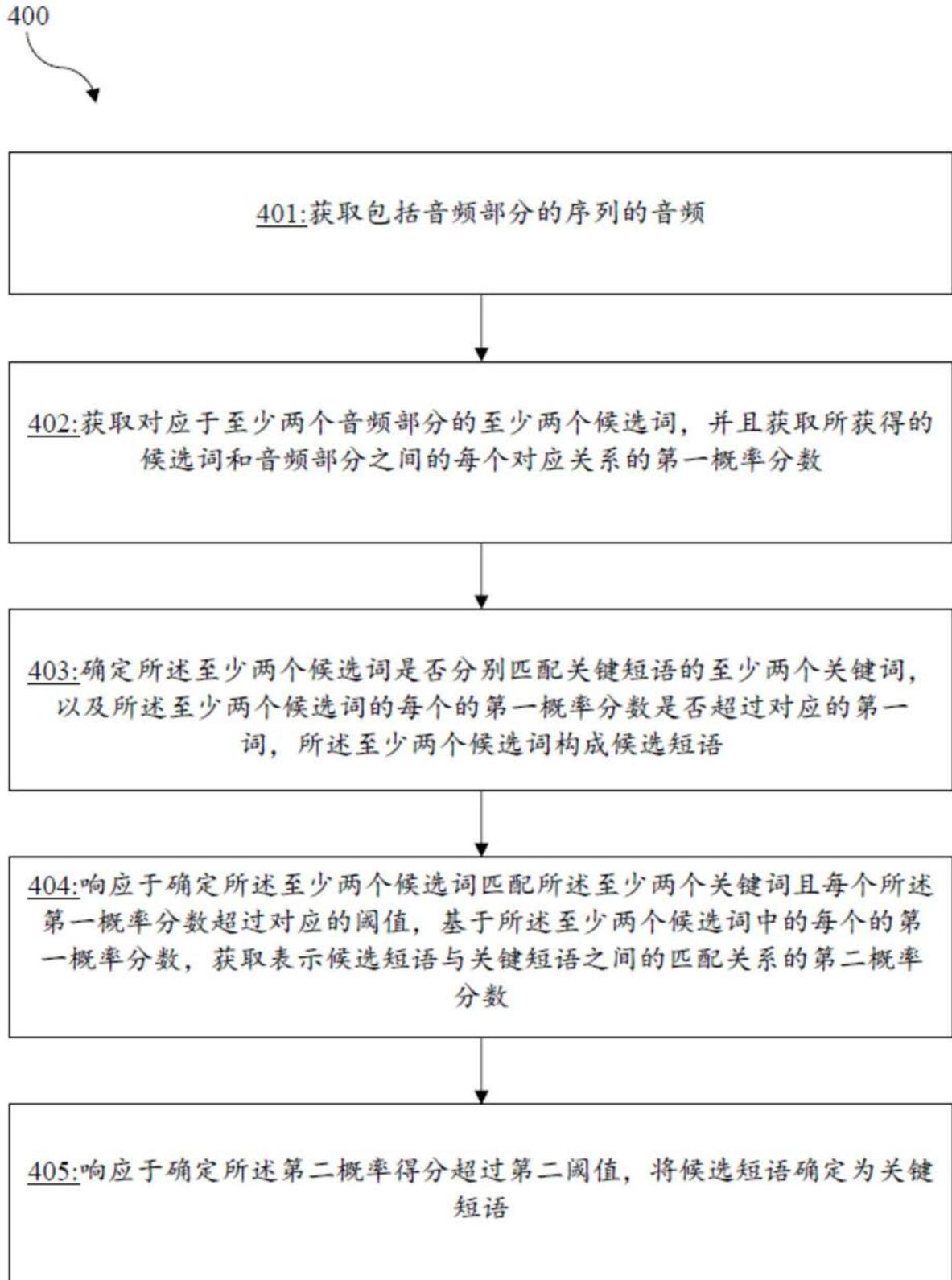


图4A



图4B

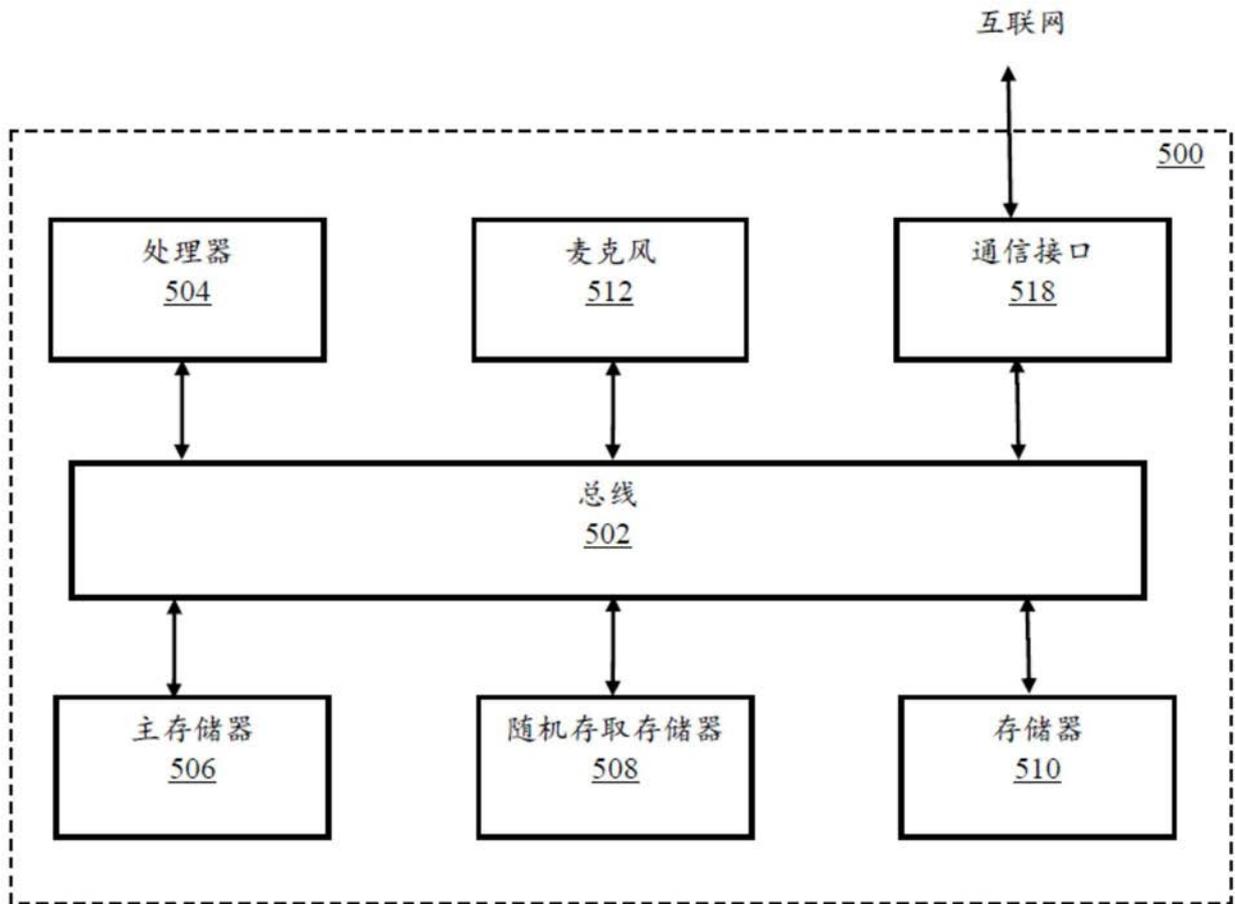


图5